Inference of influence in social networks

Matija Piškorec[1], Nino Antulov-Fantulin[1 2], Iva Miholić[3], Tomislav Šmuc[1], Mile Šikić[3 4]

Abstract: We study the problem of inference of influence in social networks, and specifically differentiating between influence which is *endogenous* and that which is *exogenous* to the network. In the case of information diffusion in online social networks, endogenous (peer) influence is specified through an explicit influence model between peers, and it corresponds to the various ways users can interact and influence each other in online social networks, for example sharing or evaluating content generated by other users. On the other hand, exogenous (external) influence is specified as acting uniformly on all users regardless of the current state of their peers, and it corresponds to interaction which is not part of the online network, for example online news sources that independently share the same content. We define a likelihood function which can include wide range of peer influence models as well as external influence, and optimize it numerically to find maximum likelihood parameters. We evaluate our methodology on simulated activation cascades using several common models of peer influence - susceptible-infected (SI), exponential decay and logistic threshold model. We also perform inference on two large Facebook networks of 10175 and 6202 users where activation cascade is an act of registration to an online political survey application which happened during a period of one week. In addition to estimates of peer and external influence in network, our methodology is also able to characterize activation of each individual user as being peer or externally driven, and to identify most influential users.

Problem formulation: Our model is conceptually similar to the unified model of social influence [2] which was shown to be generalization of many popular influence models, including complex contagion model [4], independent cascade model [3] and generalized threshold model [3]. At each time step all activated users attempt to activate all of their peers in network with certain probability. Also, there is an external influence which acts equally on all users at a given time step, although it may change over time. Activation times of each individual user are known, and together they form an *activation cascade*. Our problem is then the following: Given particular form of peer influence along with activation cascade and a network of users, infer parameters of peer and external influence while assuming that parameters of peer influence stay constant throughout the period while external influence may change in time. Similar attempts exist in literature, including peer and authority model [1] which, however, requires explicit modeling of *authorities* responsible for external influence.

Methodology: We define a likelihood function which explicitly includes both peer influence and external influence for each of the users activated in a specific time interval. By definition, external influence is equal for all users in a specific time window, while peer influence is dependent on the state of user's peers. Given some functional forms for peer and external influence $p_{peer}^{(i)}$ and $p_{ext}^{(i)}$ at time window t for each user i we define log-likelihood function $\log \mathcal{L}$:

$$\log \mathcal{L}(D; p_{peer}, p_{ext}) = \sum_{i \in \text{activated at} t} \log(1 - (1 - p_{peer}^{(i)})(1 - p_{ext}^{(i)})) + \sum_{i \in \text{nonactivated at} t} \log((1 - p_{peer}^{(i)})(1 - p_{ext}^{(i)}))$$

As an example, we test two peer influence models: 1) *susceptible-infected (SI)* model, where a_i is the number of activated friends of user i and p_{SI} is parameter of peer influence:

$$p_{peer}^{(i)} = 1 - \prod_{j \in \text{activated}} (1 - p_{SI}) = 1 - (1 - p_{SI})^{a_i}$$

and 2) *exponential decay* model, where t_j is the time of activation of user j and (p_0, λ) are parameters of peer influence:

$$p_{peer}^{(i)}(t) = 1 - \prod_{j \in \text{activated}} (1 - p_0 e^{-\lambda(t - t_j)})$$

[1]Laboratory for Information Systems, Division of Electronics, Ruđer Bošković Institute, Croatia
[2]Computational Social Science, ETH Zurich, Switzerland
[3]Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
[4]Bioinformatics Institute, A*STAR, Singapore, Republic of Singapore

Log-likelihood is optimized numerically by alternatively performing two optimization steps, similar as in [5]: (i) fixing parameters of peer influence and optimizing parameters of external influence for each time window and (ii) fixing parameters of external influence and optimizing parameters of peer influence for the whole time period.

Experiments on simulated activation cascades: We tested our methodology on simulated activation cascades generated using susceptible-infected and exponential decay peer influence models, as well as external influence which we model to approximate three distinct peaks in media attention, as shown in Figure 1.

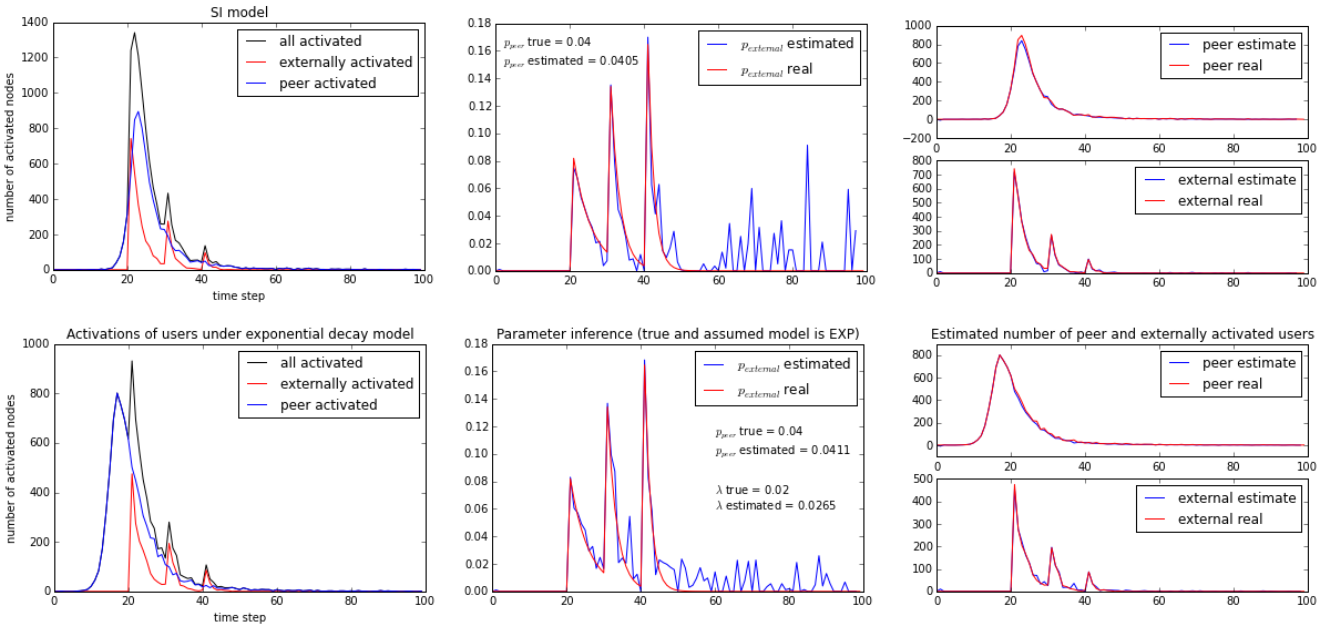

Figure 1: Inference of peer and external influence on simulated activation cascade, using susceptible-infected (above) and exponentially decaying (bellow) model of peer influence. External influence changes over time and has three distinct peaks which decay in time (middle). Our methodology successfully infers parameters of both peer models and external influence (middle) as well as absolute number of users activated due to one or the other (right).

Experiments on Facebook activation cascades: We collected two large datasets consisting of Facebook users who registered on two distinct political survey applications one week prior to the actual voting. Application used Facebook API to collect information on Facebook friendships between users, resulting in two large online social networks consisting of 10175 and 6202 users. As true motivations of users are not known, we applied our methodology on these activation cascades to obtain estimates of peer and external influence in network. Our methodology also allows us to characterize activation of each individual user as being peer or externally driven, and to identify most influential users.

[1] Anagnostopoulos, A., Brova, G. & Terzi, E. Peer and Authority Pressure in Information-Propagation Models. In *Machine Learning and Knowledge Discovery in Databases*, 76-91 (Springer Berlin Heidelberg, 2011).

[2] Srivastava, A., Chelmis, C. & Prasanna, V.K. The unified model of social influence and its application in influence maximization. In *Social Network Analysis and Mining*, 1-15 (2015).

[3] Kempe, D., Kleinberg, J. & Tardos, E. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137-146 (2003).

[4] Karsai, M., Iñiguez, G., Kaski, K. & Kertész, J. Complex contagion process in spreading of online innovation. In *Journal of The Royal Society Interface*, 11, 101 (2014).

[5] Myers, S.A., Zhu, C. & Leskovec, J. Information diffusion and external influence in networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 33-41 (2012).